MAS364/61006: Assignment 1

You can find formulae for named distributions and conjugate pairs in Appendix A of the lecture notes.

Solutions and discussion are written in blue. Some common pitfalls are indicated in teal. A sample mark scheme is given in red, with each mark placed after the statement/deduction for which the mark would be given. As usual, mathematically correct solutions that follow a different method would be marked analogously.

Marks are given for [A]*ccuracy*, [J]*ustification, and* [M]*ethod.*

⋆⋆ 1. (a) Let $X \sim \text{Pareto}(2,1)$. Find the probability density function of $X|_{\{X\geq 5\}}$.

(b) Let (X,Y) be continuous random variables with p.d.f. satisfying

$$f_{X,Y}(x,y) \propto \begin{cases} ye^{-\lambda y}x^{y-1} & \text{if } x \in (0,1) \text{ and } y \in (0,\infty), \\ 0 & \text{otherwise} \end{cases}$$

i. Find the distribution of Y.

ii. Find the distribution of $X|_{\{Y=y\}}$, for $y \in (0,\infty)$.

Solution.

(a) We have $\mathbb{P}[X|_{\{X\geq 5\}} \in [5,\infty)] = 1$ [1M] and for $A \subseteq [5,\infty)$ we have

$$\mathbb{P}[X|_{\{X\geq 5\}} \in A] = \frac{\mathbb{P}[X\in A]}{\mathbb{P}[X\geq 5]} = \frac{\int_A 2x^{-3}\,dx}{\int_5^\infty 2x^{-3}\,dx} = \frac{\int_A 2x^{-3}\,dx}{5^{-2}} = \int_A 50x^{-3}\,dx$$

[1M] It follows that

$$f_{X|_{\{X\geq 5\}}}(x) = \begin{cases} 50x^{-3} & \text{for } x \geq 5 \\ 0 & \text{otherwise.} \end{cases}$$

[2A] (The question does not ask to name it, but this is the Pareto$(5,2)$ distribution.)

Pitfall: Don't forget to specify the region on which the p.d.f. is non-zero.

(b) i. Recall that $f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)\,dx$. Hence, for $y \in (0,\infty)$ we have

$$f_Y(y) \propto \int_0^1 ye^{-\lambda y}x^{y-1}\,dx = ye^{-\lambda y}\left[\frac{x^y}{y}\right]_{x=0}^1 = e^{-\lambda y}$$

[1M] and $f_Y(y) = 0$ for $y \leq 0$. We recognize $Y \sim \text{Exp}(\lambda)$. [1A]
In particular this tells us that normalizing constant above is λ, so

$$f_Y(y) = \begin{cases} \lambda e^{-\lambda y} & \text{for } y \in (0,\infty), \\ 0 & \text{otherwise.} \end{cases}$$

ii. Here are two different ways to solve this question.

For the first way, we use that $f_{X|_{\{Y=y\}}}(y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$ [1M] and compute that

$$f_{X|_{\{Y=y\}}}(x) \propto \begin{cases} \frac{y e^{-\lambda y} x^{y-1}}{\lambda e^{-\lambda y}} & \text{for } x \in (0,1), \\ 0 & \text{otherwise,} \end{cases} = \begin{cases} y x^{y-1} & \text{for } x \in (0,1), \\ 0 & \text{otherwise.} \end{cases}$$

We recognize this p.d.f. as $X|_{\{Y=y\}} \sim \text{Beta}(y,1)$. [1A]

Alternatively: we could add a factor λ into $f_{X,Y}(x,y)$ and then we obtain

$$\begin{aligned} \mathbb{P}[X \in A, Y \in B] &\propto \int_B \int_A \lambda e^{-\lambda y} y x^{y-1} \, dx \, dy = \int_B \left(\int_A y x^{y-1} \, dx \right) \lambda e^{-\lambda y} \, dy \\ &\propto \int_B \mathbb{P}[\text{Beta}(y,1) \in A] f_Y(y) \, dy \end{aligned}$$

which matches the form of a Bayesian model (X,Y) with model family $(\text{Beta}(1,y))_{y \in (0,\infty)}$ and prior $Y \sim \text{Exp}(\lambda)$. Hence, the normalizing constant above must be 1. For this model we know that $X|_{\{Y=y\}} \sim \text{Beta}(1,y)$.

⋆ 2. In ancient Rome, the *denarius* was a commonly used coin. In this question we are interested in the weight of a denarius coin, which is variable because they were manufactured by hand. A further complication is that, as Roman craftsmanship gradually improved, smaller and smaller versions of the same coin were made.

An analysis of the production methods has suggested that we should model the weight of a single coin, in grams, as $\text{N}(\theta, 0.5^2)$. An archaeological dig finds 3 denarii coins, with weights (in grams)

$$x = (3.68, 3.92, 3.85).$$

The archaeologists believe these coins are from the era 37-14 B.C. shortly after the death of Julius Ceaser (in 44 B.C.). They are certain that Julius Ceaser set the official weight of a denarius at 3.9g, but they don't know how closely this was followed, in practice. After some consultation, we decide on a prior $\Theta \sim \text{N}(3.9, 1.2^2)$.

(a) Write down a suitable model family M_θ, which leads to a Bayesian model (X, Θ) applicable to the data x.

(b) Use the Normal-Normal conjugate pair to write down the distribution of the posterior $\Theta|_{\{X=x\}}$.

(c) *This part should be done on a computer and is not for handing in. You may it helpful to use the code provided in Excercise **2.1** (or **4.1**) of the lecture notes.*

Use R or Python to plot the prior and posterior density functions from (b) on the same axes.

(d) Suppose that, in an alternate reality, we had found the data x but had not been able to discuss the choice of prior with the archaeologists. Instead of the prior above we chose $\Theta \sim \text{N}(7.5, 4.5^2)$ – based, say, on the rough weights of various modern coins, which are usually between 3g and 12g. Find the resulting posterior $\Theta|_{\{X=x\}}$ and comment on how this compares to the results in (b).

Solution.

(a) $M_\theta \sim N(\theta, 0.5^2)^{\otimes 3}$ [1A]

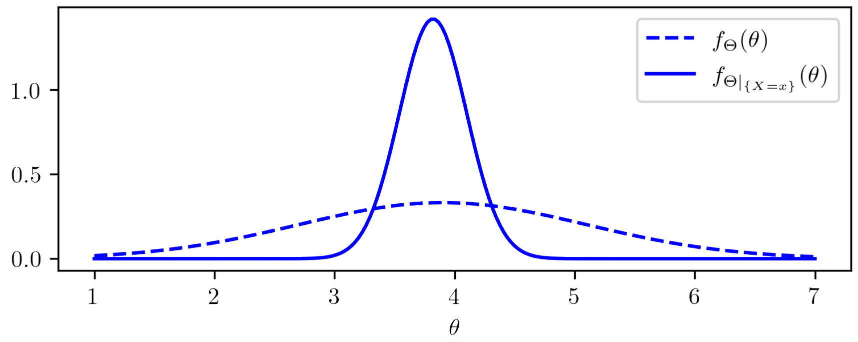
Pitfall: We need the $\otimes 3$ because we have three (independent) datapoints.

(b) Our data satisfies $\sum_1^3 x_i = 11.45$, our prior is $\Theta \sim \mathrm{N}(3.9, 1.2^2)$, and we have fixed $\sigma^2 = 0.5$ in our model. Putting these into the update equations for the Normal-Normal conjugate pair gives posterior

$$\Theta|_{\{X=x\}} \sim \mathrm{N}\left(\frac{\frac{1}{0.5^2}(11.45) + \frac{3.9}{1.2^2}}{\frac{3}{0.5^2} + \frac{1}{1.2^2}}, \frac{1}{\frac{3}{0.5^2} + \frac{1}{1.2^2}}\right) \approx \mathrm{N}(3.82, 0.0788).$$

where we have rounded the parameters in the final step. [1M + 2A]

(c) You should obtain the following graph: [2A]



(d) We have the same setup as in (b), except that now the prior is $\Theta \sim \mathrm{N}(7.5, 4.5^2)$, resulting in

$$\Theta|_{\{X=x\}} \sim \mathrm{N}\left(\frac{\frac{1}{0.5^2}(11.45) + \frac{7.5}{4.5^2}}{\frac{3}{0.5^2} + \frac{1}{1.2^2}}, \frac{1}{\frac{3}{0.5^2} + \frac{1}{4.5^2}}\right) \approx \mathrm{N}(3.83, 0.0830).$$

[1M + 2A]

In (b) we obtained a slightly different mean and a lower variance in the posterior distribution. We started out with a prior distribution in (b) that has a lower mean and was more concentrated about its mean, and used the same data in both cases, so it makes sense that the posterior distribution also has a (slightly) lower mean and is more concentrated in (b). [1J]

However, the posteriors obtained in both cases are very similar. Our choice of prior has not had much effect here. [1J]

⋆⋆ 3. In this question we use the parametrization of the Geometric distribution from the reference sheet, that is

$$\mathbb{P}[\mathrm{Geometric}(\theta) = k] = \begin{cases} \theta(1-\theta)^k & \text{for } k \in \mathbb{N}_0 = \{0, 1, 2, \ldots\}, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Let $n \in \mathbb{N}$. Let (X, Θ) be a Bayesian model with model family $M_\theta \sim \mathrm{Geometric}(\theta)^{\otimes n}$, where $\theta \in [0,1]$, and prior $\Theta \sim \mathrm{Beta}(a, b)$, where $a, b \in (0, \infty)$.

Let $x = (x_1, \ldots, x_n) \in (\mathbb{N}_0)^n$. Show that the posterior distribution is

$$\Theta|_{\{X=x\}} \sim \mathrm{Beta}\left(a + n, b + \sum_{i=1}^n x_i\right).$$

In this question you should justify your calculations ***without*** *using the reference sheet of conjugate pairs.*

(b) The probability of a wild oyster containing a pearl is around one in ten thousand. This probability can be greatly increased in farmed oysters with various techniques, but it is still an uncertain process.

An oyster farmer opens oysters one by one. We use a Geometric(θ) distribution to model the number of times the farmer opens an oyster and does not find a pearl, up to (but not including) the time at which the first pearl is found. We decide to use the prior $\Theta \sim \text{Beta}(1, 10)$. The farmer has repeated this experiment 10 times and obtained the data

$$x = (25,\ 38,\ 11,\ 60,\ 23,\ 29,\ 4,\ 28,\ 61,\ 28). \qquad (\star)$$

For this data $\sum_{i=1}^{10} x_i = 307$.

i. What is the range R and the parameter space Π of the discrete Bayesian model specified above?

ii. How many oysters have been opened, in total, to collect the data $(\star)$?

iii. Use part (a) to find the posterior distribution $\Theta|_{\{X=x\}}$, using the data supplied by the oyster farmer. What is the value of $\mathbb{E}[\Theta|_{\{X=x\}}]$ and how does this compare to our prior?

iv. Suppose that, instead, we choose to model the probability of finding a pearl inside a *single* oyster as Bernoulli(θ).
Let (Y, Θ) be a Bayesian model with model family $\text{Bernoulli}(\theta)^{\otimes 317}$, with the same prior $\Theta \sim \text{Beta}(1, 10)$. Use the reference sheet of conjugate pairs to find the posterior parameters for this model, given the data represented in $(\star)$.

v. You should obtain the same posterior distribution in parts iii and iv. Suggest why has this happened.

Solution.

(a) Using Bayes rule (in the case of discrete data) we have [1M]

$$\begin{aligned}
f_{\Theta|_{\{X=x\}}}(\theta) &\propto \mathbb{P}[\text{Geometric}(\theta)^{\otimes n} = x] f_{\text{Beta}(a,b)}(x) \\
&\propto \left(\prod_{i=1}^{n} \theta(1-\theta)^{x_i}\right)\left(\frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}\right) \\
&\propto \left(\theta^n \theta^{\sum_1^n x_i}\right)\left(\theta^{a-1}(1-\theta)^{b-1}\right) \\
&\propto \theta^{a+n-1}(1-\theta)^{b+\sum_1^n x_i - 1}.
\end{aligned}$$

for $\theta \in (0,1)$, and $f_{\Theta|_{\{X=x\}}}(\theta) = 0$ otherwise. [2M+2A]

We recognize $\Theta|_{\{X=x\}} \sim \text{Beta}(a+n, b+\sum_1^n x_i)$ as required.

(b) i. $\Pi = (0,1)$ (or $[0,1]$ will do) and $R = \{0,1,2,\ldots\}^{10}$. [2A]

ii. 317, [1A] of which 10 contained pearls and 307 did not.

iii. We apply part (a) with $n = 10$ and $\sum_1^n x_i = 307$. We obtain $\Theta|_{\{X=x\}} \sim \text{Beta}(1+10, 10+307) \sim \text{Beta}(11, 317)$. [2A]
We have $\mathbb{E}[\Theta] = \frac{1}{1+10} \approx 0.09$ and $\mathbb{E}[\Theta|_{\{X=x\}}] = \frac{11}{11+317} \approx 0.03$. Compared to our prior, this posterior distribution suggests that (on average) many fewer oysters contain pearls than we originally believed. [1J]

iv. We apply a Bayesian model (Y, Θ) with $n = 317$, model family $\text{Bernoulli}(\theta)^{\otimes 317}$ and prior $\Theta \sim \text{Beta}(1, 10)$. Our data from $(\star)$ is a vector $y \in \{0,1\}^{317}$, such that $y = (y_i)_{i=1}^{317}$ contains $k = \sum_1^{317} y_i = 10$ ones (i.e. successes) and $n - k = 307$ zeros (i.e. failures). [2M]
We obtain $\Theta|_{\{X=x\}} \sim \text{Beta}(1 + k, 10 + n - k) \sim \text{Beta}(11, 317)$. [1A]

v. If $(X_i)_{i\in\mathbb{N}}$ is a sequence of i.i.d. $\text{Bernoulli}(\theta)$ random variables, then the $\text{Geometric}(\theta)$ distribution counts number of $i = 0, 1, 2, \ldots$ for which we have $X_i = 0$, up until the first 1. [1J] We made the same transformation to the data between parts iii and iv, when we used a vector containing 307 zeros and 10 ones in place of $(\star)$. For this reason, our two models are really the same model, with different formats for the data. [1J]

Total marks: 35